Taylor & Francis
Taylor & Francis Group

Land-cover classification of China: integrated analysis of AVHRR imagery and geophysical data

J. Y. LIU†#, D. F. ZHUANG†, D. LUO‡* and X. XIAO§†

†Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China
‡Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, 100101, China
§Complex Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

Abstract. Over last two decades, numerous studies have used remotely sensed data from the Advanced Very High Resolution Radiometer (AVHRR) sensors to map land use and land cover at large spatial scales, but achieved only limited success. In this paper, we employed an approach that combines both AVHRR images and geophysical datasets (e.g. climate, elevation). Three geophysical data-sets are used in this study: annual mean temperature, annual precipitation, and elevation. We first divide China into nine bio-climatic regions, using the long-term mean climate data. For each of nine regions, the three geophysical data layers are stacked together with AVHRR data and AVHRR-derived vegetation index (Normalized Difference Vegetation Index) data, and the resultant multi-source datasets were then analysed to generate land-cover maps for individual regions, using supervised classification algorithms. The nine land-cover maps for individual regions were assembled together for China. The existing land-cover dataset derived from Landsat Thematic Mapper (TM) images was used to assess the accuracy of the classification that is based on AVHRR and geophysical data. Accuracy of individual regions varies from 73% to 89%, with an overall accuracy of 81% for China. The results showed that the methodology used in this study is, in general, feasible for large-scale land-cover mapping in China.

1. Introduction

Accurate and timely information on land use and land cover at the regional scale is needed for biogeochemical, hydrological and climate modelling. Over the last two decades, numerous studies have used remotely sensed images from meteorological satellites (e.g. Advanced Very High Resolution Radiometer (AVHRR) sensors) to map land use and land cover at large spatial scales. The AVHRR sensors provide daily observations of the Earth at 1–4 km spatial resolutions. The Normalized Difference Vegetation Index (NDVI), which is calculated either from the digital

---

*Present address: Department of Geography, University of Maryland, Maryland, USA.
#e-mail address: liujy@igsnrr.ac.cn

number or reflectance values of Channel 1 (red band) and Channel 2 (near-infrared band) of the AVHRR sensors, is related to vegetation condition and phenology (Goward *et al.* 1985, Lloyd 1990). Multi-temporal AVHRR data have been used to generate land-cover maps for various regions in the world, e.g. Africa (Tucker *et al.* 1985), North America (Goward *et al.* 1985, Loveland *et al.* 1991, Zhu and Evans 1994) and South America (Stone *et al.* 1994). To meet the need of global land-cover data for global change research (IGBP 1990, 1992), multi-temporal AVHRR data were also used to produce global land-cover maps (Defries and Townshend 1994, Defries *et al.* 1995, Loveland *et al.* 2000, Hansen *et al.* 2000).

The AVHRR NDVI-based approach has made substantial contribution to land-cover characterization at regional and global scales, however there is still room for further improvements. The relatively low accuracy is due to both the coarse spatial and spectral resolutions of AVHRR data, and the confusions between some land-cover types (Scepan 1999, Scepan *et al.* 1999). When performing land-cover classification, problems exist in discerning land-cover types with similar phenology (Townshend *et al.* 1991), as some land-cover types share similar spectral reflectance characteristics.

At regional to global scales the spatial distribution of vegetation is closely related to environmental conditions, including elevation, climate (temperature and precipitation) and soils. Rapid development and wide applications of Geographical Information System (GIS) technology in natural resources and environment have generated many kinds of geographical datasets that are related to land-cover types. Remotely sensed data are also an important information source for GIS in natural resources management (Trotter 1991). A few studies have demonstrated that incorporation of ancillary data embodied within a GIS can aid interpretation and labelling of spectral clusters and improve land-cover classification. When classifying a Landsat Multi-Spectral Scanner (MSS) image for Olympic National Park, Washington, Cibula and Nyquist (1987) used terrain data and climatological data (temperature and precipitation), and the number of land-cover types that could be differentiated increased from 9 to 21 classes, and the overall accuracy of the classification increased to 91.7%. Liu *et al.* (1998) used a GIS-embodied model that described observed relationships between land cover, slope and elevation, and significantly improved the vegetation classification of a Landsat Thematic Mapper (TM) image of Helan Mountain in China. Brown *et al.* (1993) used multi-source data (coarse resolution satellite images and ancillary data) for global land-cover characterization. Ancillary data (digital elevation, temperature, precipitation and frost-free period) were used for post-classification refinement of AVHRR data, and an improved map of land-cover types for the conterminous USA was generated (Brown *et al.* 1993). Giri and Shrestha (1996) used terrain data and 1 km AVHRR images to map vegetation distribution in Bangladesh and an overall classification accuracy of 75% was reached. Liu *et al.* (1998) used geophysical data (precipitation, temperature, elevation) to generate a geo-environmental image by principal component analysis (PCA), which was then overlaid on an AVHRR image for a supervised vegetation classification of Northeast China. The new method substantially improved the land-cover classification in Northeast China, in comparison to the conventional classification method that was based only on AVHRR image data (Liu *et al.* 1998).

In this study we combined coarse resolution 1997 AVHRR data and ancillary geophysical data (elevation, mean annual precipitation and mean annual temperature) to map land cover in China. The geophysical datasets are organized and

processed within a GIS. Fine-resolution land-cover maps derived from Landsat TM data were used for image interpretation and validation. The objectives of the study are twofold: (1) to quantitatively assess the potential of the multi-source data (remote sensing/GIS) approach for land-cover classification at the continental scale; and (2) to generate an accurate and updated land-cover map of China at 1 km spatial resolution.

2. Description of AVHRR imagery and geophysical datasets

2.1. *AVHRR data*

The Chinese National Bureau of Meteorology has ground receiving stations for AVHRR sensors and is responsible for acquisition and pre-processing of daily AVHRR data for China. Routine pre-processing of AVHRR data include radiometric correction, geometric correction and atmospheric correction. Monthly and seasonal composites of AVHRR data are also generated and provided to users, based on maximum values of NDVI. Four seasonal composites of AVHRR data in 1997 were acquired for this study: Spring (February to April 1997), Summer (May to July 1997), Fall (August to October 1997) and Winter (November and December 1997, and January 1998). All four seasonal composites include all the channels of AVHRR data and NDVI at 1 km spatial resolution. The images have a dimension of 4280 rows and 5502 columns, covering all the territory of China, and are in the Albers Equal Area Map Projection.

2.2. *Geophysical datasets*

A number of geophysical datasets were collected and organized for this study, including elevation, climate, fine-resolution land-use and land-cover maps, and administrative boundary maps. The digital elevation (contour) dataset was originally digitized and refined from topographic maps at 1:1 000 000 scale by the National Geographical Information Center of National Survey and Mapping Bureau. The vector-format elevation data were converted into raster (grid) format at 1 km spatial resolution (figure 1), using Arc/Info GIS software. For climate data we collected daily weather records of 361 meteorology stations in China over the period of 1959–1990 from the Chinese Bureau of Meteorology, and calculated mean annual temperature (figure 2), annual precipitation (figure 3), cumulated temperature (base 0°C, 10°C), and humidity index. A point coverage was generated for all the 361 weather stations, using Arc/Info GIS software. Climate data from the weather stations were then interpolated into a 500 m by 500 m grid, and the temperature fields were further adjusted using the Digital Elevation Model (DEM; 1:1 000 000), following the assumption that temperature decreases by 0.6°C when elevation increases by 100 m. Because of the nature of the source data, these climate data represent continental climate conditions and does not represent local or microclimate conditions. Elevation and climate data are input datasets for supervised classification, while the other geophysical datasets were used for masking, interpretation and labelling of classification clusters. All the geophysical datasets are in Albers Equal Area Map Projection (with the same projection parameters as the AVHRR dataset).

To select training signatures of various land-cover types for AVHRR imagery, we use (1) the Vegetation Map of China at the scale of 1:4 000 000, which was produced by the Institute of Botany, Chinese Academy of Sciences; and (2) the National Land Cover Dataset (NLCD-96) at 1:100 000 scale (Liu and Buhaeosier 2000). The Vegetation Map of China, which was based on field surveys and satellite
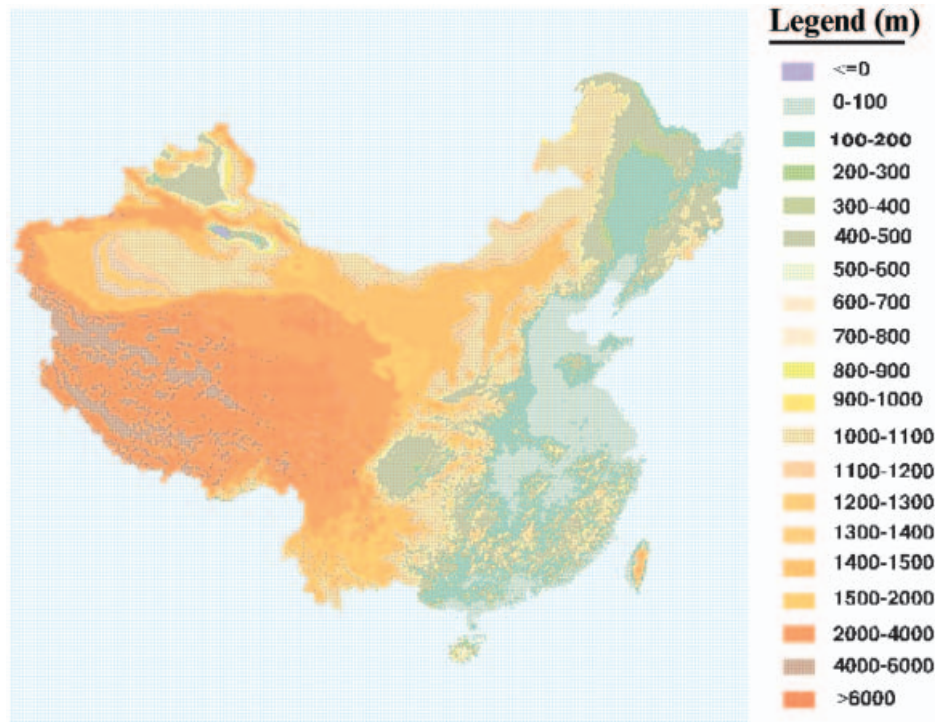
Figure 1.   Digital elevation model of China.

images in the early 1980s, was used primarily to define training signatures of forest types. The NLCD-96 dataset is composed of 25 land-use and land-cover types and was used as the main reference for signature training and consistency checking. The NLCD-96 dataset was developed through the National Land Cover Project that was organized by the Chinese Academy of Sciences with participation of over 100 scientists from eight research institutions. Over 500 Landsat TM images acquired in 1995 and 1996 were used to generate 1:100 000 scale maps of land cover in China (Liu and Buheaosier 2000). It is temporally and spatially adequate for signature training and accuracy assessment. It should also be noted that the NLCD-96 data classes represent a mix of land use and land cover. Most 'natural' landscape units (e.g. wet land and forest classes) represent land cover, whereas those heavily impacted by anthropogenic activities (e.g. urban and agricultural classes) represent land use. The primary focus of this study is to map land cover and, thus, the NLCD-96 classes are not always compatible with the target land-cover classes in this study (table 1).

When performing land-cover classification using AVHRR data, cities could not be uniquely identified because of complex mixtures of surface conditions within a 1 km urban pixel. City distribution information was extracted from the NLCD-96 dataset and rasterized to a binary city mask. After classification of AVHRR imagery, it was added into the classification result. Maps of province and country political boundaries at a scale of 1:1 000 000 in 1995 from the National Survey and Mapping Bureau were used to summarize the classification results by province and country.
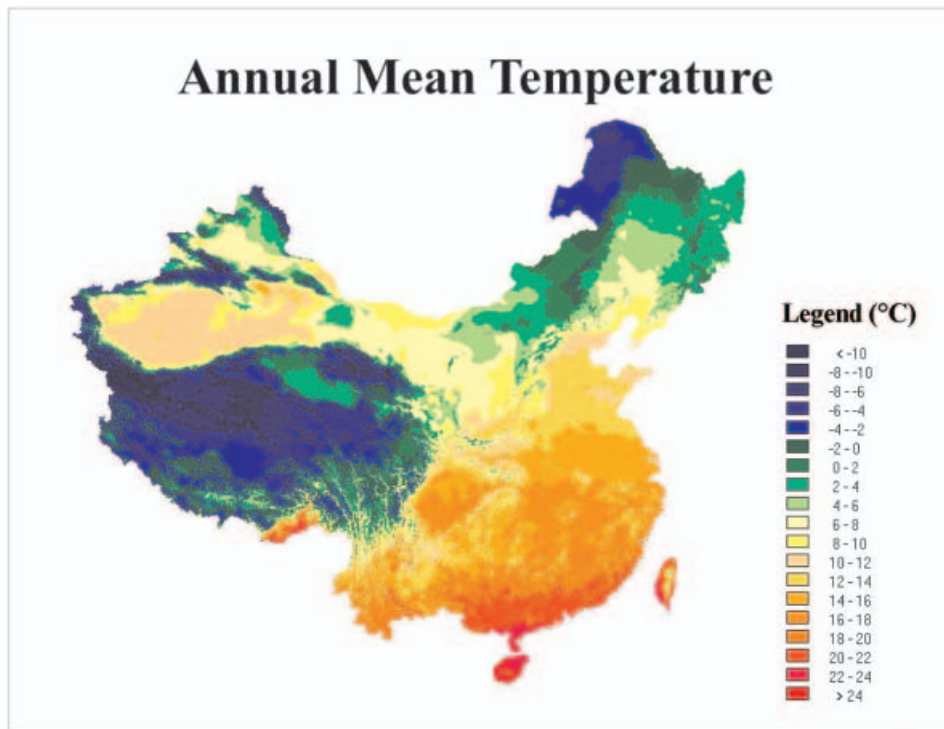
Figure 2.   Annual mean temperature of China.

3.   Methods

Land-cover mapping at continental to global scales needs to address unique and difficult issues in data analysis: (1) large spatial and temporal variations in climate, terrain, vegetation, and soil; and (2) large volumes of data, which require a high capacity of computing resources (Loveland *et al.* 1991). In this study for China, we used the following approach to address the first issue: to partition the continental-scale study area into several smaller regions, and carry out classifications for individual regions using different data layers that better describe land-cover characteristics in a region, and finally mosaic resultant maps of individual regions into one continental map of land cover. This approach involves three major phases (figure 4). The first phase was to develop a regionalization scheme for China, based on climate data. The second phase was to conduct a supervised classification for each of the climatic regions. Finally, the third phase was to evaluate classification results by region and create a mosaic product of land cover in China using the classification results from different climatic regions.

At large spatial scales, vegetation growth and distribution are primarily associated with climatic factors such as temperature and moisture. Aridity index (Gao 1996) and cumulative temperature (base 0, 5, or 10° C) over a year are well-known indicators of bio-climatic conditions and regions. The gridded long-term mean climate fields (see §2) were used to calculate aridity index (Gao 1996) and cumulative temperature (base 10° C) in order to generate a map of climatic regionalization for China. According to the criteria of temperature and moisture conditions (table 2), nine climatic regions were defined for China (figure 5). The advantages of the climatic
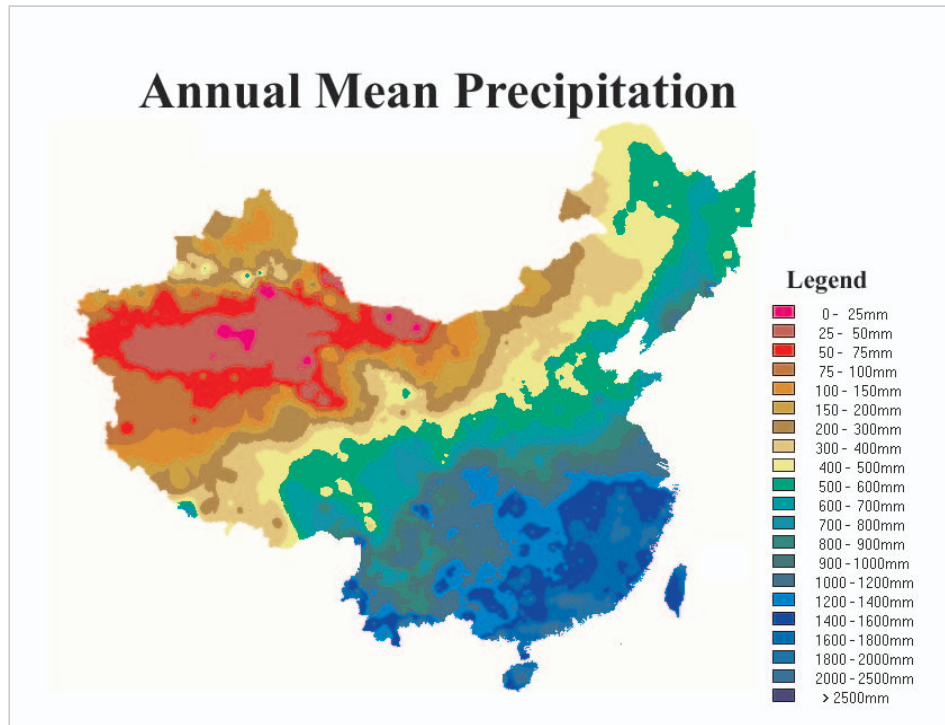
Figure 3.   Annual mean precipitation of China.

regionalization approach include (1) defining a relatively homogeneous region and having the potential of reducing classification confusion in a region; (2) selecting appropriate temporal AVHRR images according to unique phenological characteristics of different regions; and (3) delineating training signatures for supervised classification with a higher confidence. Therefore, for all nine climatic regions of China (figure 5), different sets of temporal NDVI images, AVHRR data, and climate station data were selected (table 2). For Northeast China and North China, four seasons of NDVI data were selected to calculate annual accumulated NDVI, in an attempt to maximize the differences among various land-cover types. For the other seven regions, only the images for the plant growing season (i.e. summer images) were used.

PCA is a useful tool for analysis of multi-spectral remote sensing images (Roger 1996). It transforms a set of correlated spectral bands of image data into an equivalent set of uncorrelated components of data. The main advantage of PCA is that the transformed components can be ordered by decreasing variances. Typically, for remote sensing data, just a few of the low-order components have large variance and contain useful information, and the other high-order components look like noise. In an earlier study (Liu *et al.* 1998), it was found that the first principal component derived from AVHRR data contains over 80% of the information of AVHRR images, and can be used as a representative band for vegetation interpretation. The image of the first principal component (PC1) was generated to compress multi-spectral imagery (Liu *et al.* 1998). In this study, PCA was used to process AVHRR data for the nine regions of China and the resultant PC1 images were used for analysis.

Table 1. Description of the land-cover classification system used in this study and the estimates of individual land-cover types for China from supervised classification of AVHRR and geophysical datasets.

| Land-cover types	Symbol	Description	Area (km²)
Evergreen needleleaf forest	C1	Lands dominated by needleleaf trees which remain green all year.	379 258
Deciduous needleleaf forest	C2	Lands dominated by seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	128 864
Evergreen broadleaf forest	C3	Lands dominated by broadleaf trees which remain green all year.	145 254
Deciduous broadleaf forest	C4	Lands dominated by seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	232 845
Mixed forest	C5	Lands dominated by tree communities with interspersed mixtures or mosaics of the four kinds of forest cover type above. None of the forest types exceeds 50%.	58 029
Shrub	C6	Lands with woody vegetation less than 2 m.	1 080 855
Dense grassland	C7	Lands with herbaceous types. The canopy cover is >60%.	928 933
Moderate dense grassland	C8	Lands with herbaceous types. The canopy cover is 20–60%.	508 940
Sparse grassland	C9	Lands with herbaceous types. The canopy cover is 5–20%.	958 085
Farmland	C10	Lands covered with temporary crops followed by harvest and a bare soil period (e.g. single and multiple cropping systems).	2 077 498
Wetland	C11	Lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas.	56 809
City	C12	Land covered by buildings and other man-made structures.	36 915
Water body	C13	Oceans, seas, lakes, reservoirs and rivers.	124 235
Ice and snow	C14	Lands under ice and/or snow and never have more than 5% vegetated cover in a decade.	79 397
Harsh desert	C15	Lands expose soil and stone and never have more than 5% vegetated cover.	1 232 057
Sandy desert	C16	Lands expose sand and never have more than 5% vegetated cover.	548 045
Bare rock	C17	Lands expose rock and never have more than 5% vegetated cover.	485 507
Forest in Tibet region	C18	Lands in Tibet region and dominated by trees including conifer, broadleaf and shrub.	436 129
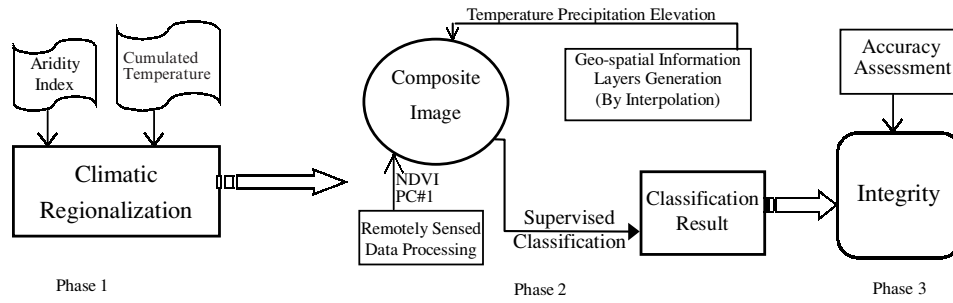
Figure 4. Processing flow phase 1: climatic regionalization of China using maps of cumulated temperature (base 10°C) and aridity index. Phase 2: perform supervised classification for each sub-region using the composite image derived remote sensing data and geo-spatial information. Phase 3: accuracy assessment and classification results integrity.

Table 2. The criteria for defining bio-climatic regions of China, and AVHRR images that were used for NDVI and PCA analysis for each bio-climatic region.

| Region | Code | Criteria for bio-climatic region | | NDVI | PCA |
		Temperature	Aridity index		
South China	SC	$T > 6500$	$A_r < 1.0$	Su	Su
East China	EC	$5100 < T < 6500$	$A_r < 1.0$	Su	Su
Central China	CC	$3300 < T < 5100$	$A_r < 1.0$	Sp, Su, F, W	Su
North China	NC	$3300 < T < 5100$	$1.0 < A_r < 1.5$	Sp, Su, F, W	Su
ShanJin Area	SJ	$1700 < T < 4500$	$1.5 < A_r < 4.0$	Sp, Su, F, W	Su
Inner Mongolia	IM	$1700 < T < 4500$	$1.0 < A_r < 2.0$	Su	Su
Northwest China	NWC	$1700 < T < 4500$	$A_r > 4.0$	Su	Su
Northeast China	NEC	$1700 < T < 3300$	$A_r < 1.5$	Sp, Su, F, W	Su
Tibet	TB			Su	Su

Seasonal AVHRR composites: Sp—spring, Su—summer, F—fall, and W—winter, T—cumulative temperature over a year (base 10° C); A_r—aridity index, calculated using the following equation: $A_r = 0.16T/r$, where r is the precipitation in millimetres (Gao 1996).

For each of the nine regions, five data layers (NDVI, PC1, precipitation, temperature and elevation) are stacked together and used for supervised classification. The integrated image/geophysical datasets contain both spectral and geophysical information of the land surface. Supervised clustering algorithms and maximum likelihood classifications were applied to the datasets of the nine regions to generate classification clusters. The NLCD-96 dataset and published maps, including Vegetation Map of China (Hou *et al.* 1982) and Atlas of Grassland Resources of China (Sue 1993), were used for selection of training signatures in the AVHRR images, and interpretation and labelling of spectral clusters in relation to land-cover types.

In North China, Northeast China and East China, where there are diverse land-cover types and similar spectral characteristics between grassland and some forests, the binary decision tree method was applied to reduce classification confusion. Forest distribution information was extracted from the NLCD-96 dataset using GIS, and a binary mask of forest and non-forest was generated. The mask (forest, non-forest) was applied separately, while performing classification of the integrated image/geophysical dataset using the maximum likelihood algorithm (figure 6).
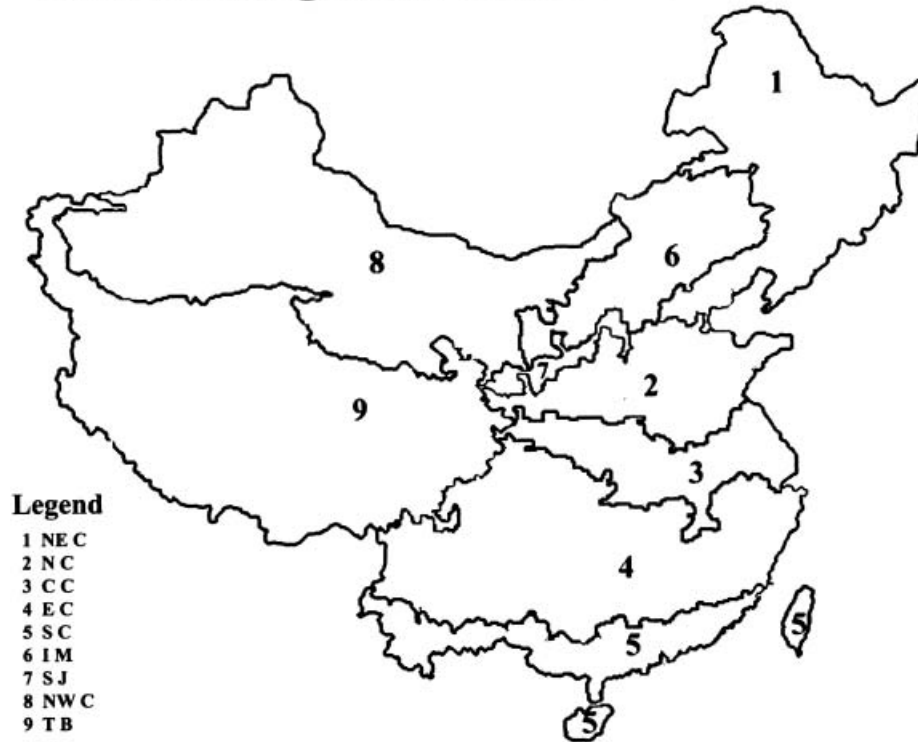
Climatic Regionalization of China

Figure 5. Climatic regionalization of China.
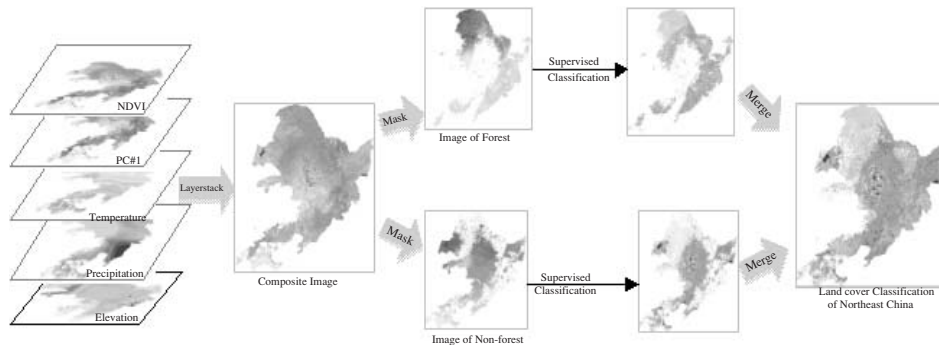
Figure 6. Land-cover classification of Northeast China using binary decision tree.

In this study, the Tibet Plateau was given additional consideration, due to its unique environmental characteristics. The Tibet Plateau is referred to as 'the third pole', which implies not only its high elevation, but also its extreme coldness. Furthermore, the interaction of moisture and heat in the plateau, which contributes greatly to the complex distribution of vegetation, are also incomparable to other mountain areas that span a relatively small range. There are few meteorological stations in the plateau, and climate data are inadequate to describe the precipitation

Figure 7.   Land-cover classification of China.
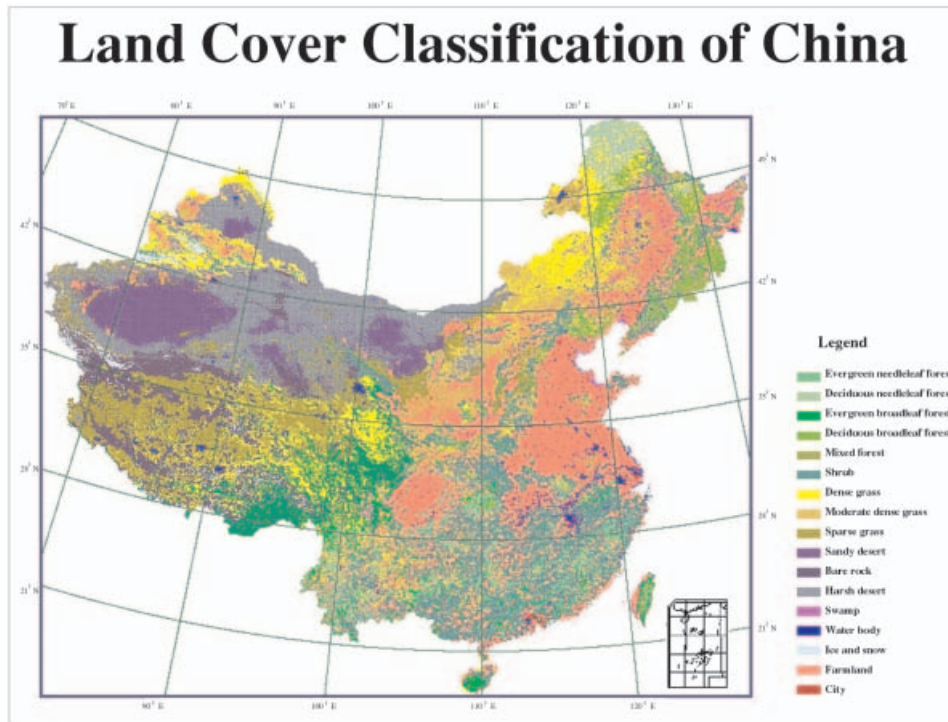
and temperature conditions. Therefore, we selected only three bands (NDVI, PC1 and DEM) to conduct the classification. In addition, most of the forests in the plateau are distributed in narrow river valleys, making it difficult to discriminate different forest types at 1 km resolution. In this study, we only assign one forest cover type for land-cover classification in the plateau due to these difficulties in forest identification and discrimination.

After interpretation and labelling of classification clusters in relation to land-cover types for individual regions, classification accuracy assessment by regions was conducted using the NLCD-96 dataset. The modified classification maps from individual regions were assembled to generate a mosaic product of land cover for the all of China. It is possible that some pixels along the adjacent boundaries in two climate regions are one land-cover type (e.g. forest), but were assigned to two different types of land cover according to the classification results for these two climate regions, which would result in a 'sudden' but 'false' change in land-cover distribution in the mosaic product. Land-cover mismatch along the boundaries between climate regions is an inevitable problem that needs to be carefully addressed. Adjustment should be done along the boundaries to make the land-cover allocation more reasonable. In this study, the majority of pixels along the boundaries matched fairly well. There is only a small amount of pixels that have land-cover mismatch along the boundaries of climate regions, mostly attributed to the supervised classification and the interspersed and cursive area boundary derived from climatic regionalization. For those pixels that have mismatch in land cover along the adjacent boundaries,

the NLCD-96 dataset was used for adjustment of land-cover types between the adjacent boundaries.

4. Results and discussion

The basic vegetation distribution in China shown in the classification result could be summed as Forests in north-eastern China, south-eastern China and part of the Tibet region; Grassland in the Inner-Mongolia region, and Southwest China; farmland in the eastern part of the country; and large areas of sparsely vegetated land and bare ground in Northwest China (figure 7). Total areas for all classes are shown in table 1.

Assessment of land-cover classification accuracy is a complex issue. The coarse resolution of AVHRR data leads to development of classes based commonly on land-cover mosaics rather than on homogeneous landscapes. Accessibility of consistent site data for verification is also a limitation. An additional complication is the fact that the existing land-cover information is not based on well-defined categories (Loveland *et al.* 1991). As a result, verification was limited to comparisons with other datasets or maps. In this study, the land-cover data layers were compared with the NLCD-96 dataset and vegetation map. Together, these data sources provide users with useful information regarding the quality of the final land-cover product. A sample of 200 AVHRR 1 km pixels was randomly selected within each of the climate regions. The geo-locations of those pixels were determined on the appropriate reference map (NLCD-96 and vegetation map) for comparison. Class values of those pixels from the AVHRR-derived final classification were obtained and compared with the NLCD-96 dataset and vegetation map.

The assessment results (tables 3 and 4) showed that there is reasonably good agreement between the land-cover dataset and the reference data. The overall accuracy of the land-cover classification was 80.6%. Some land-cover change likely occurred between the dates of the AVHRR images and the dates of the reference data (NLCD-96 and vegetation map in 1980s). It is difficult to ascertain the impact of actual land-cover change on these values in this study. In addition, any errors in the reference data will decrease the overall agreement.

The result from classification of AVHRR/geophysical data shows that land-cover distribution patterns are reasonably consistent with other existing maps. Of the 17

Table 3. Accuracy assessment of land-cover classification for the nine bio-climatic regions in China. It includes the overall accuracy and Kappa coefficient for all land-cover types in a region.

Region	Code	Overall accuracy (%)	Kappa (%)
South China	SC	83.7	78.4
East China	EC	79.5	68.7
Central China	CC	88.6	82.0
North China	NC	73.3	64.6
ShanJin area	SJ	75.2	65.1
Inner Mongolia	IM	77.6	70.8
Northwest China	NWC	77.1	65.1
Northeast China	NEC	84.8	75.3
Tibet	TB	84.3	76.2
Overall accuracy		80.6	

Table 4. Accuracy assessment of different land-cover types.

Symbol	Index	Symbol of each region								
		SC	EC	CC	NC	NEC	SJ	NWC	TB	IM
C1	E_o	88.89	70.00	77.78	80.00	–	–	–	–	100.00
	E_c	58.54	65.63	70.00	90.24	–	–	–	–	100.00
	κ	0.4892	0.578	0.686	0.8155	–	–	–	–	1.0000
C2	E_o	–	–	–	–	93.75	–	83.33	–	–
	E_c	–	–	–	–	83.30	–	100.00	–	–
	κ	–	–	–	–	0.7431	–	0.8002	–	–
C3	E_o	52.78	53.33	–	–	–	–	–	–	–
	E_c	70.37	44.44	–	–	–	–	–	–	–
	κ	0.6659	0.4032	–	–	–	–	–	–	–
C4	E_o	–	–	42.86	100.00	69.23	100.00	–	–	50.00
	E_c	–	–	60.00	75.00	69.23	66.67	–	–	100.00
	κ	–	–	0.5556	0.746	0.6725	0.6634	–	–	1.0000
C5	E_o	40.00	50.00	–	–	72.22	–	–	–	–
	E_c	100.00	66.67	–	–	81.25	–	–	–	–
	κ	1.00	0.6604	–	–	0.7030	–	–	–	–
C6	E_o	66.67	81.16	89.19	72.34	75.00	75.00	55.56	–	77.27
	E_c	86.39	83.58	86.84	75.56	78.95	93.75	100.00	–	70.83
	κ	0.8215	0.7593	0.8389	0.6758	0.7679	0.8872	0.5455	–	0.7069
C7	E_o	–	30.43	66.67	57.14	83.33	40.00	84.62	84.72	80.00
	E_c	–	77.87	66.67	66.67	76.92	66.67	64.71	87.14	78.57
	κ	–	0.7514	0.6617	0.654	0.7373	0.6567	0.832	0.7238	0.7044
C8	E_o	100.00	75.00	50.00	71.43	66.67	69.49	66.67	100.00	76.67
	E_c	75.00	54.55	87.50	55.56	72.72	65.08	46.51	66.67	74.19
	κ	0.7067	0.4556	0.8657	0.3691	0.7112	0.5319	0.7078	0.6473	0.6313
C9	E_o	–	–	–	70.00	100.00	–	76.92	74.36	40.00
	E_c	–	–	–	77.78	100.00	–	41.67	78.38	100.00
	κ	–	–	–	0.7655	1.0000	–	0.6269	0.6506	1.0000

Table 4. (*Continued*).

Symbol	Index	Symbol of each region								
		SC	EC	CC	NC	NEC	SJ	NWC	TB	IM
C10	E_o	93.65	87.93	95.88	85.06	95.08	83.87	73.33	100.00	90.24
	E_c	75.64	76.10	93.91	90.24	95.08	79.59	84.62	71.43	86.05
	κ	0.6035	0.6678	0.8618	0.8155	0.9300	0.628	0.7117	0.7072	0.8245
C11	E_o	–	–	–	100.00	75.00	–	–	–	50.00
	E_c	–	–	–	83.33	60.00	–	–	–	50.00
	κ	–	–	–	0.8289	0.5924	–	–	–	0.4949
C12	E_o	100.00	100.00	100.00	100.00	100.00	100.00	100.00	–	100.00
	E_c	100.00	100.00	100.00	100.00	100.00	100.00	100.00	–	100.00
	κ	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	–	1.0000
C13	E_o	100.00	100.00	100.00	100.00	71.43	–	100.00	–	100.00
	E_c	100.00	100.00	80.95	100.00	83.33	–	100.00	–	100.00
	κ	1.0000	1.0000	0.792	1.0000	0.8227	–	1.0000	–	1.0000
C14	E_o	–	–	–	–	–	–	100.00	100.00	–
	E_c	–	–	–	–	–	–	66.67	80.00	–
	κ	–	–	–	–	–	–	0.6434	0.7961	–
C15	E_o	–	–	–	–	–	75.00	84.00	56.25	50.00
	E_c	–	–	–	–	–	75.00	100.00	90.00	50.00
	κ	–	–	–	–	–	0.745	0.7989	0.8863	0.4792
C16	E_o	–	–	–	–	–	–	90.91	–	–
	E_c	–	–	–	–	–	–	100.00	–	–
	κ	–	–	–	–	–	–	0.8015	–	–
C17	E_o	–	–	–	–	–	–	100.00	97.62	–
	E_c	–	–	–	–	–	–	76.92	91.11	–
	κ	–	–	–	–	–	–	0.7010	0.8086	–
C18	E_o	–	–	–	–	–	–	–	84.62	–
	E_c	–	–	–	–	–	–	–	75.86	–
	κ	–	–	–	–	–	–	–	0.7241	–

E_o is the error of omission expressed as a percentage; E_c is the error of commission expressed as a percentage; κ is the Kappa statistics. The meaning of each symbol in the first column is listed in table 1.

classes, water bodies, wetland, ice/snow, and desert are well identified, while other land-cover types are dependent upon the climate regions. Farmland in Northeast China, North China, Central China, Northwest China and Inner Mongolia regions are clearly identifiable. Shrubland is well identified in the Central China, East China and ShanJin regions. Satisfactory classification results of grassland in Inner Mongolia and Northeast China were also obtained. Confusion of land-cover types varies in different climate regions. In the east part of China, where there is adequate precipitation, confusion occurs mainly among different forest types such as conifer, broadleaf, mixed forest and shrub. In the semi-arid and arid areas, confusion tends to occur between grasslands of differing densities, deserts and wetlands. Confusion between farmlands and grasslands seems to occur in every climate region where both of these two land-cover types co-exist. This is most serious in the ShanJin region, where farmland and grassland mix in an interspersed distribution, leading to spectral and temporal confusion in the classification.

In those climatic regions that are dominated by relatively large and homogeneous landscapes (e.g. Central China, North China, Inner Mongolia, Northwest China), land-cover types were relatively well identified, indicating that the strategy used is acceptable in those regions for land-cover mapping. However, in spatially complex areas such as East China and South China, the result of the classification was not as good as was expected. Different vegetation types in these regions exhibited both spectral and seasonal similarities. In addition, because of the homogeneous climatic environment in these regions, geophysical data cannot effectively resolve the confusion among natural land-cover types. Therefore, some other kinds of geophysical data may be needed for improved land-cover classifications for those climate regions.

To further demonstrate the effectiveness of our approach of combining AVHRR data and geophysical data, a supervised classification was performed in the Inner Mongolia region using an AVHRR image alone. There is serious confusion among farmland and different kinds of grassland. Its overall classification accuracy is only 58%, much lower than the accuracy (77.8%) from the approach that used both remotely sensed data and geophysical information (table 3). Inclusion of geophysical data makes a great contribution to the reduction of confusion in the land-cover classification. Even though the comparison was only done in one climate region, the accuracy difference is great enough to encourage further research in this area. Comparisons in other climate regions should be part of future work to testify the effectiveness of geophysical data in confusion reduction for different regions.

This preliminary evaluation indicates that the procedures used are generally acceptable. This research has also illuminated many issues that remain to be addressed. For example, the outcome of the NDVI-based classification was clearly influenced by the weather during 1997. This influence is obvious on grassland with different amounts of coverage. It is well known that growth of grass is closely correlated to moisture conditions. Precipitation variation will directly lead to changes in grass density, which will likely affect grassland classification of that year. Whether there was climatic anomaly in 1997, and its specific effects on the classification are uncertain and remain to be investigated.

It is also likely that the land-cover classification was affected by the quality of the geophysical data. In the land-cover classification, problems existed in topographically complex areas such as the Yungui Plateau in East China region and Taihang Mountain in North China region where the classification accuracy is below 60%. As these regions cover a large latitudinal distance, elevation influence on climatic

conditions varies. Because of the generalized nature in the source data, geophysical data used here is related to continental climate conditions, not local or microclimate conditions, which made it inadequate for representing vegetation patterns related to elevation zones in irregular terrain. To achieve acceptable classification results, more detailed geophysical data are necessary.

Selection of geophysical data is another issue that needs to be discussed. Precipitation, temperature and elevation are used in this study, however they are not the only factors that influence land-cover distribution. In spatially complex regions such as East China and South China, where precipitation, temperature and elevation have a limited role of reducing confusion of land-cover classification, some other geophysical factors including slope, aspect and soil texture should be taken into consideration.

5. Conclusion

Geophysical data, used in conjunction with remotely sensed data, can improve the presentation of driving factors (climate, soil, etc.) and the spatial patterns of vegetation distribution. This helps to improve the interpretation of images and contributes greatly to increased accuracy and higher efficiency of remote sensing classifications. Climatic regionalization was applied in this work to simplify the study area into relatively small and environmentally homogeneous sub-areas for reducing the confusion among land-cover types belonging to different regions. The result of the classification indicates that this is a feasible way for large-scale land-cover characterization in China. For further improvement of the classification, the following work is required: (1) acquisition of new and more detailed geophysical data; (2) improved analysis techniques in data exploration and visualization; and (3) investigation in relationships among geophysical data and land-cover types.

References

BROWN, J. F., LOVELAND, T. R., and MERCHANT, J. W., 1993, Using multi-source data in global land-cover characterization: concept, requirements, and methods. *Photogrammetric Engineering and Remote Sensing*, **6**, 977–987.

CIBULA, W. G., and NYQUIST, M. O., 1987, Use of topographic and climatological models in a geographical data base to improve Landsat MSS classification for Olympic National Park. *Photogrammetric Engineering and Remote Sensing*, **1**, 67–75.

DEFRIES, R., and TOWNSHEND, J., 1994, NDVI-derived land classifications at a global scale. *International Journal of Remote Sensing*, **17**, 3567–3586.

DEFRIES, R., HANSEN, M., and TOWNSHEND, J., 1995, Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data. *Remote Sensing of Environment*, **54**, 209–222.

GAO, G., (ed.), 1996, *Introduction to Climatology* (Beijing: Meteorology Press) [in Chinese].

GIRI, G.., and SHRESTHA, S., 1996, Land-cover mapping and monitoring from NOAA AVHRR data in Bangladesh. *International Journal of Remote Sensing*, **14**, 2749–2759.

Goward, S. N., Tucker, C. J., and Dye, D. G., 1985, North American vegetation patterns observed with the NOAA-7 Advanced Very High Resolution Radiometer. *Vegetatio*, **64**, 3–14.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., and Sohlberg R., 2000, Global Land Cover Classification at 1 km resolution using a classification tree approach. *International Journal of Remote Sensing*, **21**, 1331–1364.

Hou, X., Sun, S., Zhang, J., He, M., Wang, Y., Kong, D., and Wang, S., 1982, *Vegetation Map of the People's Republic of China* (Beijing: Map Publisher of China).

IGBP, 1992, *Global Change and Terrestrial Biosystems: The Operational Plan*, IGBP Global Change Report No. 20, (Stockholm: IGBP).

International Geosphere–Biosphere Programme, 1990, Global Change: Report No. 12 (Stockholm: IGBP Secretariat).

Liu, J., Zhuang, D., and Ling, Y., 1998, Vegetation integrated classification and mapping using remote sensing and GIS techniques in Northeast China. *Journal of Remote Sensing*, **4**, 285–291.

Liu, J., and Buheaosier, 2000, Study on spatial–temporal feature of modern land-use change in China: using remote sensing technique. *Quaternary Sciences*, **3**, 229–239 [in Chinese with English abstract].

Liu, W., Gong, J., and Fang, H., 1998, Information extraction from GIS data base and its applications in vegetation classification. *Journal of Remote Sensing*, **3**, 235–239.

Lloyd, D., 1990, A phenological classification of terrestrial vegetation using shortwave vegetation index imagery. *International Journal of Remote Sensing*, **12**, 2269–2279.

Loveland, T. R., Merchant, J. W., and Brown, J. F., 1991, Development of a Land-Cover Characteristics Database for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*, **11**, 1453–1463.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W., 2000, Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. *International Journal of Remote Sensing*, **21**, 1303–1330.

Roger, R. E., 1996, Principal components transform with simple, automatic noise adjustment. *International Journal of Remote Sensing*, **14**, 2719–2727.

Scepan, J., Menz, G., and Hansen, M. C. 1999, The DISCover Validation Image Interpretation Process. *Photogrammetric Engineering and Remote Sensing*, **65**, 1075–1081.

Scepan, J., 1999, thematic validation of high-resolution global land-cover data sets. *Photogrammetric Engineering and Remote Sensing*, **65**, 1051–1060.

Stone, T., Schlesinger, P., Houghton, R. A., and Woodwell, G. M., 1994, A map of the vegetation of South America based on satellite imagery. *Photogrammetric Engineering and Remote Sensing*, **60**, 541–551.

Sue, D. (chief editor), 1993, *Atlas of Grassland Resources of China* (1:1 000 000 scale) (Beijing: Press of Mapping).

Townshend, J. R. G., Justice, C. O., Li, W., Gurney, C., and McManus, J., 1991, Global land-cover classification by remote sensing: present capabilities and future possibilities. *Remote Sensing of Environment*, **2**, 243–255.

Trotter, C. M., 1991, Remotely-sensed data as an information source for geographical information systems in natural resources management—a review. *International Journal of Geographical Information Science*, **2**, 225–239.

Tucker, C. J., Townshend, J. R. G., and Goff, T. E., 1985, African land cover classification using satellite data. *Science*, **4685**, 369–375.

Zhu, Z., and Evans, D. L., 1994, U.S. forest types and predicted percent forest cover from AVHRR data. *Photogrammetric Engineering and Remote Sensing*, **5**, 525–531.